Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction

Edward J. O'Brien, Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, Bernhard Ø. Palsson

Corresponding author: Bernhard Ø. Palsson, University of California San Diego

Review timeline:

Submission date:	22 April 2013
Editorial Decision:	17 June 2013
Revision received:	18 July 2013
Editorial Decision:	03 September 2013
Revision received:	04 September 2013
Accepted:	05 September 2013

Editors: Maria Polychronidou, Thomas Lemberger

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 17 June 2013

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who agreed to evaluate your manuscript. As you will see from the reports below, the reviewers acknowledge that your work is addressing a potentially interesting topic. However, they raise a series of concerns and make suggestions for modifications, which should be carefully addressed in a revision of the manuscript.

Without repeating all the points listed below, several of the reviewers' comments refer to the need to provide additional validations in order to better demonstrate the predictive power of the model. Additionally, the referees require explanations and/or clarifications regarding the employed methodology as well as assumptions and adjustments mentioned throughout the manuscript. Finally, Reviewer #3 recommends cautious interpretation of the presented results in order to avoid overstatements.

REFEREE REPORTS:

Reviewer #1:

In this study O'brien et al. reconstruct the first Metabolic-Expression (ME)-model for E.coli. This model is an extension of the E.coli genome-scale metabolic model to include the process of gene expression as well. The ME-model is used to predict several cellular phenotypes such as the cell's maximum growth rate, substrate uptake and by product secretion rates, central carbon metabolic flux

and gene product expression levels.
The reconstruction of new ME-model is with no doubt of much value, and this work has a great potential in making a contribution to the field of cellular modeling. However, additional explanations and validations are required to assess the model's predictive power. Below we give specific suggestions for further strengthening the paper. Assuming those can be adequately addressed, this paper is bound to make an important contribution and be of much interest to the wide readership of Molecular Systems Biology.

General Comments:
1. The authors report 91.3% accuracy in predicting gene essentiality. Which cutoff has been used to define an essential gene? Are the results sensitive to the specific choice of this value?
2. In Supplementary Table S4 the authors compare the gene essentiality prediction of the M-model to those obtained by the ME-model. It would be beneficial and more informative if the authors could provide the exact accuracy obtained by the M-model alone, together with precision and recall obtained by the two models.
3. The authors set the model's parameters based on previously published studies. However, in some cases the authors explicitly say that a parameter is a rough approximation. Can the authors perform a sensitivity analysis and examine how the model's predictions are affected by these values?
4. In figures 1C, 1E-F the authors plot the in silico vs. experimental results, showing they are in agreement. This result should be supported by a correlation and P-value.
5. In the section "Growth regions under varying nutrient availability" the authors report that the ME-model's performance is similar to that of Adadi et al. First, the authors should explicitly report the correlation obtained by the ME-model in the main text. Secondly, according to Supplementary Table 5, the method of Adadi et al. outperforms the ME-model when the latter is with `solvent accessible', and obtains similar results when the ME-model is without `solvent accessible'. The authors should explain what is solvent accessible (?) and why it affects the model's performance.
6. Similar to comment #4, results presented as plots alone in Figure 2C-D should be accompanied with correlations and P-values, and reported in the main text.
7. When comparing the model's flux predictions to experimentally measured flux rate: (1) What is the correlation between the predicted and measured flux rate? (2) What is the performance of the M-model alone in this case?
8. When comparing the model's gene expression predictions to the experimentally measured ones, the authors report that the model correctly predicts changes in the right direction for three clusters of genes. (1) What is the overall prediction accuracy of the model when looking at the entire gene set? (2) Does the model predict gene expression changes that also quantitatively agree with experimental data (that is, beyond changes in the right direction)?
9. The authors use hypergeometric test to identify TFs more associated than expected by chance with genes change during the Janusian shift. (1) The authors should further work to clarify the writing of this part as it is unclear. (2) Was the p-value corrected for multiple hypotheses? (3) In figure 5E the authors should quantify the agreement of their data with experimental measurements instead of the just showing them in a plot.
10. Generally, whenever flux rates and gene expression are predicted, the authors should perform a sampling analysis to further support their finding since the flux solution is not unique,.

Minor comments:
1. Page 4: "Michaelis-Menten-type rate law (Figure 1C)" Michaelis-Menten-type rate law (Figure 1D).
2. Please revise the writing of assumption (2) in the first paragraph of page 5 as it is not clear.
3. Page 6: "the cell increases its growth rate though" the cell increases its growth rate through.

Reviewer #2 :
The authors present a genome-scale model for E. coli that is the result of an integration of a published metabolic reaction network from year 2011 and a network that describes gene expression and synthesis of functional macromolecules published in 2009. The topology of the model is explored and the model itself is used to analyze the changes in growth rates upon various limiting conditions. The model is further used to explain how protein availability can explain growth rate limitation even when substrate availability is not the limiting factor. Furthermore, the outputs of the model are successfully compared to various already published data on protein/gene expression.

This is an interesting paper that, although it does not provide any new biological insights on the studied systems, does provide a method that can be used to simulate growth and changes in genome scale systems. The presented method links metabolism and protein production machinery. Therefore, it provides a more accurate and extensive description than a regular genome scale metabolic model.

Previously, there have been other successful attempts to integrate these types of information. The approach showed in this manuscript could be, in principle, extended to other organisms in a more or less straightforward way. In that sense, it is a valuable contribution to the field.

Comments

In this study the model is used to predict gene essentiality and the results are provided in a supplementary table. The paper would benefit from elaborating on these results. For example, why are some genes essential in the M-Model and not essential in the ME-Model? The author should also explain what "inconsistent" means in this table.

On the derivation of constraints for molecular catalytic rates, two minimal assumptions are given to account for the changes in metabolic catalysis, namely:
(1) When the cell is nutrient-limited, protein content is maximized (at a given growth rate).
(2) This protein mass is metabolic enzymes not operating at their maximal catalytic rate.
It should be explained why these assumptions are made and some arguments (plausibility arguments) should be provided.

On page 5, 4th paragraph, the following is written:
" A recent approach in which kinetic parameters are used to bound metabolic flux in M-Models performs similarly to the ME-Model on this dataset...". The information that supports this claim is presented in a supplementary table. This assertion is based on Pearson's correlation coefficients that show this similarity, but only partially (0.4535 versus 0.4681). However this second value is only obtained upon adjusting the keff . It should be explained why this adjustment is required, otherwise this sentence becomes an overstatement.

Paragraph "Adaptive evolution gene expression data processing" from Materials and methods is unclear:
"For Figure 6B, k-means clustering (k=5).... as in (Conrad et al., 2010))."
Why is k=5 chosen? It is also unclear where I can see the two gene clusters that consistently decrease and the two gene clusters that consistently increase. Furthermore, their relationship with Figure 6B is unclear. Please clarify.

In supplementary figure 2, for both figures, the scale of the x-axis should equal the scale of the y-axis. Or a line should be drawn, indicating where the fold changes are exactly equal. Otherwise it gives the false impression that predicted and measured values are very similar, when they are not.

In figure 2B it should be indicated either in the picture itself, or in the figure legend what the green and blue areas represent? Is it as in figure 2a Batch and SNL?

In figure 4, the link between 4B and 4C should be made clearer. Either by indicating on the graphs in 4B the values of the ratios for the relative expression values at growth rates 0.93 and 0.45 h-1 as in figure 4C, or by indicating in 4C the corresponding values for the graphs of 4B (by numbering or assigning letters, for example). I think that the way Figure 4B is drawn now is a bit misleading. The chosen range for the growth rates leads to higher fold changes of the gene-enzyme pair expression than to those presented in 4C, which by the way are not too high.

In figure 5, the red and the red-brownish colors are very hard to distinguish, please change these into distinguishable colors.

It is not straightforward to find the model on http://opencobra.sourceforge.net. A more detailed explanation where exactly the model can be found would be helpful.

Typos:

Page 2, 2nd paragraph, second sentence:
"What distinguishes these models from each other is the underlying assumptions..."
should be changed to:
"What distinguishes these models from each other are the underlying assumptions..."

Page 2, 3rd paragraph, first sentence:
"Here, we construct a ME-Model..."
should be changed to:
"Here, we construct an ME-Model..."

Page 6, 4th paragraph, second sentence:
"In the Janusian region, the cell increases its growth rate though differential expression..."
should be changed to:
"In the Janusian region, the cell increases its growth rate through differential expression..."

Page 11, 4th paragraph, first sentence:
"We finally show the ability of the ME-Model to simulate systems-level mechanism underlying adaptive evolution..."
should be changed to:
"We finally show the ability of the ME-Model to simulate systems-level mechanisms underlying adaptive evolution..."

Reviewer #3 :
The present paper represents a major advance in constraints-based metabolic modeling by incorporating expression of individual genes, not just a singular biomass objective function. If the key methodological issues (itemized below) can be addressed, and the claims (esp. in title and abstract) toned down to better match the results, I am supportive of publication:
1. Title: I do not see the paper as presenting a unified model of growth since it does not deal with regulation, cell division, geometry, etc. Also, looking carefully, I am not clear that it predicts different proteomes for different nutrient limitations (e.g., does it predict high GS expression in N-limitation?).
2. Abstract: The term "molecular details" is misleading. This suggests a much more precise understanding than provided here. Also, predictions of gene expression are not validated (e.g. what happens in different nutrient limitations?)
3. General: I find the first half of the results (up to the point of the gene expression analysis) to be highly important and an excellent test of the capabilities of the model. Thereafter, the paper seems to go into a highly speculative direction that exceeds the capabilities of the model.
4. Evolution: The comments about evolution are dubious as these may be limited to strains with low pyrE expression (pyrimidine pseduoauxotrophy).
5. Fluxes: Figure 3 should also present flux changes across conditions (how well are these predicted?), and should be supported by a table that provides detailed data/model predictions. Outliers should be discussed.
6. Figure 1C: Why are translation rates other than 20 AA/s not explored? It seems a modest change could better fit the data.
7. Methods: The optimization procedure is a hugely important aspect of this work, and should not be limited to the supplement.
8. Batch optimization method: The procedure is not clear. I am an expert in the field but cannot understand it. Why is a random peptide invoked? Why are there multiple LP problems in the process? A MUCH BETTER EXPLANATION IS ESSENTIAL.
9. Nutrient-limited optimization method: The procedure is clear to me, but does it differentiate the importance of different enzymes in different nutrient limitations? It seems not. If so, explain how. If not, a clear disclaimer is needed.
10. Janusian regime definition: Cool name, and interesting concept, but the evidence that this really exists only in C-limitation is missing. Seems like it would naturally exist also in other limitations, but that this just isn't captured in the present model.

1st Revision - authors' response 18 July 2013

We have revised our manuscript (MSB-13-4501) based on the reviewers' comments and questions and addressed all issues that were raised.

Below, please find our detailed response to the comments of the reviewers.

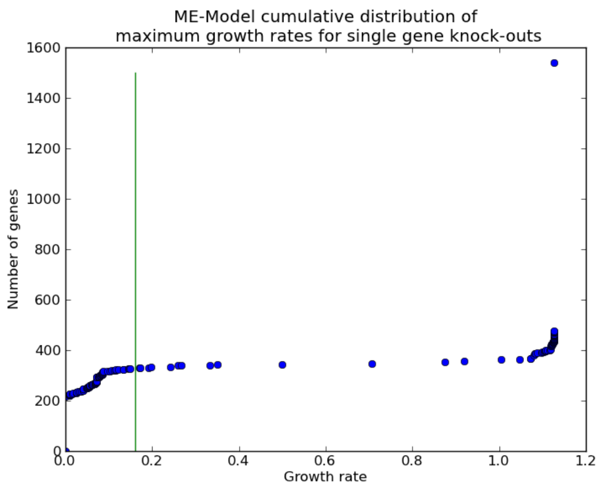
We thank the reviewers for their thoughtful comments and questions, and hope the revised manuscript will now be acceptable for publication.

Reviewer #1

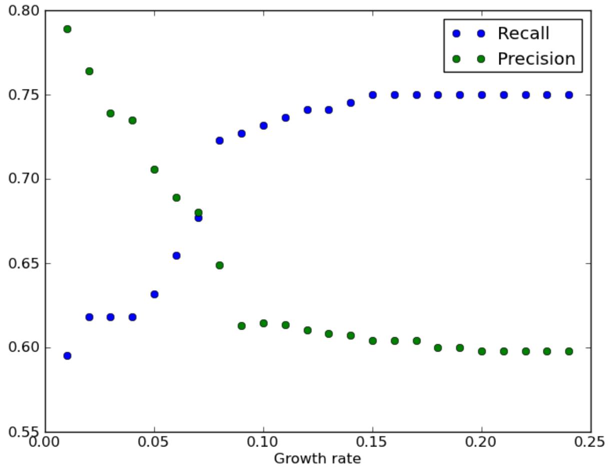
In this study O'brien et al. reconstruct the first Metabolic-Expression (ME)-model for E.coli. This model is an extension of the E.coli genome-scale metabolic model to include the process of gene expression as well. The ME-model is used to predict several cellular phenotypes such as the cell's maximum growth rate, substrate uptake and by product secretion rates, central carbon metabolic flux and gene product expression levels.
The reconstruction of new ME-model is with no doubt of much value, and this work has a great potential in making a contribution to the field of cellular modeling. However, additional explanations and validations are required to assess the model's predictive power. Below we give specific suggestions for further strengthening the paper. Assuming those can be adequately addressed, this paper is bound to make an important contribution and be of much interest to the wide readership of Molecular Systems Biology.

General Comments:
1. The authors report 91.3% accuracy in predicting gene essentiality. Which cutoff has been used to define an essential gene? Are the results sensitive to the specific choice of this value?

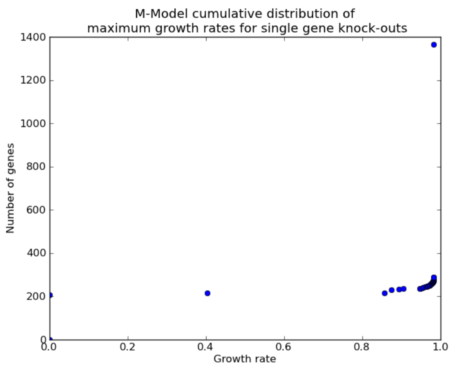
We have standardized this analysis. The results can be found in Supplementary Table 4. A gene is now considered essential for the simulated condition if deletion of the gene reduces optimal growth rate to <0.15 h^{-1} in glucose minimal media. This cutoff was derived from inspecting the cumulative distribution of maximum growth rates after single gene deletions (see below). The green line in the graph below indicates the growth rate cutoff.



As is clear from the cumulative distribution, the specific growth rate cutoff will affect the essentiality results. For this reason, we have provided the maximum growth rate after single gene deletion. For reference on the effect of the growth rate cutoff, the precision and recall as a function of cutoff are plotted below.



Compared to the growth rates in the M-Model after single gene deletion (below), the ME-Model has many more genes that result in low but non-zero growth when knocked-out (see figure below). We believe this is (at least partially) due to the multi-scale nature of the ME-Model, resulting in undetected numerical infeasibilities (i.e., the genes are actually essential, but difficult below the precision of the linear programming solver). In the ME-Model gene knockouts are accomplished through deleting the translation reaction(s) for the peptide; as these are very low flux at low growth rates, they can fall below the feasibility tolerance even for the 80 bit solver (soplex) that we use. For this reason, we provide gene expression fluxes as well from exact simplex routines available in the QSopt_ex package (Applegate, D. L., Cook, W., Dash, S. & Espinoza, D. G. Exact solutions to linear programming problems. *Operations Res. Lett.* **35**, 693–699 (2007)). Improved linear programming solvers and methods will alleviate this difficulty in the future.



2. In Supplementary Table S4 the authors compare the gene essentiality prediction of the M-model to those obtained by the ME-model. It would be beneficial and more informative if the authors could provide the exact accuracy obtained by the M-model alone, together with precision and recall obtained by the two models.

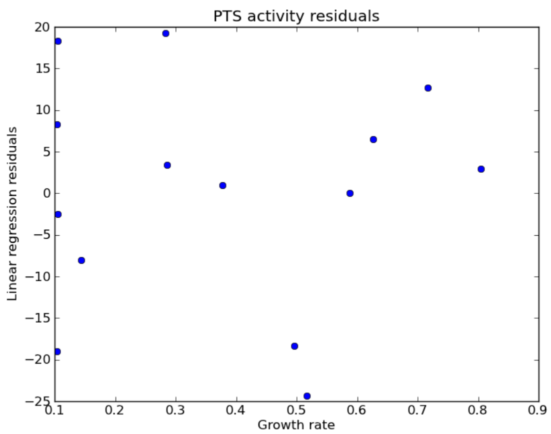
These metrics have been added to the main text.

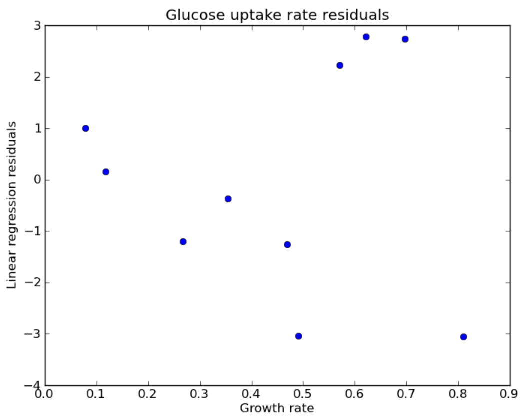
3. The authors set the model's parameters based on previously published studies. However, in some cases the authors explicitly say that a parameter is a rough approximation. Can the authors perform a sensitivity analysis and examine how the model's predictions are affected by these values?

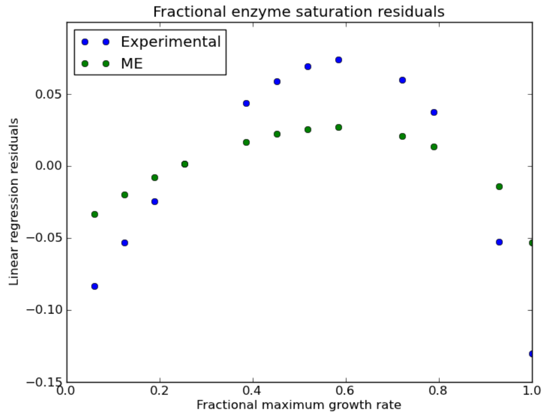
A sensitivity analysis was performed for the following parameters as they are the most uncertain parameters in the model: 1) unmodeled protein proportion of proteome (Q), and 2) median enzyme efficiency (k_{eff}). The results can be found as Supplementary Table 6. Briefly, we evaluated how changing the parameters alter optimal growth rates, substrate uptake rates, and yields across carbon sources. While the parameters affect the quantitative growth rates and substrate uptake rates (and therefore can be tuned to achieve quantitative values in future studies), the correlation with experimental data is robust.

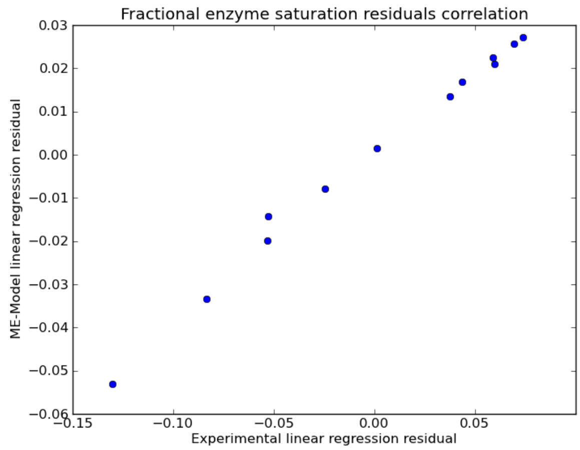
4. In figures 1C, 1E-F the authors plot the in silico vs. experimental results, showing they are in agreement. This result should be supported by a correlation and P-value.

For Figure 1C, the correlation and p-value are 0.96 and 1.24e-05, respectively. For Figures 1E and 1F, we were trying to illustrate that the ME-Model approximates the shape of the k_{eff} vs. growth rate. To support this claim (more quantitatively), we first support that the data in 1E is well-modeled by a linear fit. We have computed a two-sided hypotheses test whose null hypothesis is that the slope is zero; the two lines have p-values of 2.27e-11 and 2.61e-9. Furthermore, a plot of the residual to the linear fits (below) appears random. To show the shape of the ME-Model and the experimental k_{eff} (derived from the ratio of the linear fits in 1C) are similar, we compute the residuals to a linear fit for the lines in Figure 1F (below); the residuals have a similar (non-random) concave shape and are highly correlated (below, Pearson r=0.99, p-val=6e-12).









5. In the section "Growth regions under varying nutrient availability" the authors report that the ME-model's performance is similar to that of Adadi et al. First, the authors should explicitly report the correlation obtained by the ME-model in the main text. Secondly, according to Supplementary Table 5, the method of Adadi et al. outperforms the ME-model when the latter is with `solvent accessible', and obtains similar results when the ME-model is without `solvent accessible'. The authors should explain what is solvent accessible (?) and why it affects the model's performance.

An extensive analysis was performed to address this comment. The results can be found in the text in section "Growth regions under varying nutrient availability" and Supplementary Table 5C. 'Solvent accessible' refers to setting the enzyme k_{eff} to be dependent on the solvent accessible surface area of the enzyme (based on Miller et. al 1987, as described in Supplementary Information, Metabolic Enzymes); this was used in all simulations in the manuscript and so we now only report the results with 'solvent accessible'.

6. Similar to comment #4, results presented as plots alone in Figure 2C-D should be accompanied with correlations and P-values, and reported in the main text.

We calculated the root-mean-square deviation (RMSD) or root-mean-square error (RMSE) for the data vs. the M-Model and the data vs. the ME-Model. Lower values correspond to a better fit.

The results for Figure 2C were:

RMSE: data vs. the M-Model	=	0.40
RMSE: data vs. the ME-Model	=	0.12

The results for Figure 2D were:

RMSE: data vs. the M-Model	=	0.07
RMSE: data vs. the ME-Model	=	0.04

These values were added to the legend for Figure 2.

7. When comparing the model's flux predictions to experimentally measured flux rate: (1) What is the correlation between the predicted and measured flux rate? (2) What is the performance of the M-model alone in this case?

This correlation has been added to the main text and a comparison to M-model is in Supplementary Table 7. Additionally, there is a section in the Supplementary Information "Discussion of central carbon flux predictions" discussing the similarities/differences.

8. When comparing the model's gene expression predictions to the experimentally measured ones, the authors report that the model correctly predicts changes in the right direction for three clusters of genes. (1) What is the overall prediction accuracy of the model when looking at the entire gene set? (2) Does the model predict gene expression changes that also quantitatively agree with experimental data (that is, beyond changes in the right direction)?

This information has been added to the main text. For reference, see the sentence: "Globally, the ME-Model is not yet predictive of quantitative gene expression changes; the correlation over the entire data set is statistically significant ($p < 0.005$), but weak (Pearson's r=0.14)."

9. The authors use hypergeometric test to identify TFs more associated than expected by chance with genes change during the Janusian shift. (1) The authors should further work to clarify the writing of this part as it is unclear.

The text has been clarified.

(*2) Was the p-value corrected for multiple hypotheses?*

We have corrected the p-value for multiple hypothesis testing using the Bonferoni correction. Figures 5E and 5F were updated accordingly.

(3) In figure 5E the authors should quantify the agreement of their data with experimental measurements instead of the just showing them in a plot.

We now quantify the agreement in the main text.

10. Generally, whenever flux rates and gene expression are predicted, the authors should perform a sampling analysis to further support their finding since the flux solution is not unique.

The solutions in ME-models are more unique in M-Models (Lerman et al., 2012, Nat Commun, 3, 929). We have reported the flux variability of fluxes in central carbon metabolism in Supplementary Table 7. This shows that the solution is very constrained for these fluxes. The small variability in all reactions (which is larger for lower growth rates) is due to the allowed error in the maximum growth rate (which is required as growth optimization is a binary search instead of a single Linear Program).

Minor comments:
1. Page 4: "Michaelis-Menten-type rate law (Figure 1C)" Michaelis-Menten-type rate law (Figure 1D).
2. Please revise the writing of assumption (2) in the first paragraph of page 5 as it is not clear.

3. Page 6: "the cell increases its growth rate though" the cell increases its growth rate through.

These typos have been fixed and the assumption is now stated more clearly.

Reviewer #2

The authors present a genome-scale model for E. coli that is the result of an integration of a published metabolic reaction network from year 2011 and a network that describes gene expression and synthesis of functional macromolecules published in 2009. The topology of the model is explored and the model itself is used to analyze the changes in growth rates upon various limiting conditions. The model is further used to explain how protein availability can explain growth rate limitation even when substrate availability is not the limiting factor. Furthermore, the outputs of the model are successfully compared to various already published data on protein/gene expression.

This is an interesting paper that, although it does not provide any new biological insights on the studied systems, does provide a method that can be used to simulate growth and changes in genome scale systems. The presented method links metabolism and protein production machinery. Therefore, it provides a more accurate and extensive description than a regular genome scale metabolic model.

Previously, there have been other successful attempts to integrate these types of information. The approach showed in this manuscript could be, in principle, extended to other organisms in a more or less straightforward way. In that sense, it is a valuable contribution to the field.

Comments

In this study the model is used to predict gene essentiality and the results are provided in a supplementary table. The paper would benefit from elaborating on these results. For example, why are some genes essential in the M-Model and not essential in the ME-Model? The author should also explain what "inconsistent" means in this table.

We added more details in the text as to some of the primary differences between M-Model and ME-Model essentiality predictions (and further information in Supplementary Table 4). The definition of inconsistent has been added to Supplementary Table 4.

On the derivation of constraints for molecular catalytic rates, two minimal assumptions are given to account for the changes in metabolic catalysis, namely:
(1) When the cell is nutrient-limited, protein content is maximized (at a given growth rate).
(2) This protein mass is metabolic enzymes not operating at their maximal catalytic rate.
It should be explained why these assumptions are made and some arguments (plausibility arguments) should be provided.

Additional citations and plausibility arguments are provided in both the main text and Supplemental Information, "Optimization procedure".

On page 5, 4th paragraph, the following is written:
" A recent approach in which kinetic parameters are used to bound metabolic flux in M-Models performs similarly to the ME-Model on this dataset...". The information that supports this claim is presented in a supplementary table. This assertion is based on Pearson's correlation coefficients that show this similarity, but only partially (0.4535 versus 0.4681). However this second value is

only obtained upon adjusting the keff . It should be explained why this adjustment is required, otherwise this sentence becomes an overstatement.

An extensive analysis was performed to address this comment. The results can be found in the text in section "Growth regions under varying nutrient availability" and Supplementary Table 5.

Paragraph "Adaptive evolution gene expression data processing" from Materials and methods is unclear:
"For Figure 6B, k-means clustering (k=5).... as in (Conrad et al., 2010))."
Why is k=5 chosen? It is also unclear where I can see the two gene clusters that consistently decrease and the two gene clusters that consistently increase. Furthermore, their relationship with Figure 6B is unclear. Please clarify.

k=5 was chosen for the analysis in Conrad et al., 2010. As we did not collect this data, we deferred all data processing methods to the original authors. Conrad et al. described 5 clusters, 2 of which consistently increased and 2 of which consistently decreased (1 of the clusters did not change). These can be seen as Figure 3 in Conrad et al., 2010 (PMID: 21057108). Since we never refer to these clusters, and only the metabolic subsystems annotated in the model, we have removed these confusing references. We hope the relationship with Figure 6B is clear now.

In supplementary figure 2, for both figures, the scale of the x-axis should equal the scale of the y-axis. Or a line should be drawn, indicating where the fold changes are exactly equal. Otherwise it gives the false impression that predicted and measured values are very similar, when they are not.

The axes were updated to have the same scale.

In figure 2B it should be indicated either in the picture itself, or in the figure legend what the green and blue areas represent? Is it as in figure 2a Batch and SNL?

The figure legend has been updated accordingly.

In figure 4, the link between 4B and 4C should be made clearer. Either by indicating on the graphs in 4B the values of the ratios for the relative expression values at growth rates 0.93 and 0.45 h-1 as in figure 4C, or by indicating in 4C the corresponding values for the graphs of 4B (by numbering or assigning letters, for example). I think that the way Figure 4B is drawn now is a bit misleading. The chosen range for the growth rates leads to higher fold changes of the gene-enzyme pair expression than to those presented in 4C, which by the way are not too high.

Supplementary Figure 3 (new) implements these suggestions.

In figure 5, the red and the red-brownish colors are very hard to distinguish, please change these into distinguishable colors.

The colors have been made more distinguishable.

It is not straightforward to find the model on http://opencobra.sourceforge.net. A more detailed explanation where exactly the model can be found would be helpful.

The model was included as a "Related Manuscript File" with the previous submission (and this submission). We will post the version of the ME-Model used for this manuscript and all future updates at the URL above once the manuscript is published.

Typos:

Page 2, 2nd paragraph, second sentence:
"What distinguishes these models from each other is the underlying assumptions..."
should be changed to:
"What distinguishes these models from each other are the underlying assumptions..."

Page 2, 3rd paragraph, first sentence:
"Here, we construct a ME-Model..."
should be changed to:
"Here, we construct an ME-Model..."

Page 6, 4th paragraph, second sentence:
"In the Janusian region, the cell increases its growth rate though differential expression..."
should be changed to:
"In the Janusian region, the cell increases its growth rate through differential expression..."

Page 11, 4th paragraph, first sentence:
"We finally show the ability of the ME-Model to simulate systems-level mechanism underlying adaptive evolution..."
should be changed to:
"We finally show the ability of the ME-Model to simulate systems-level mechanisms underlying adaptive evolution..."

These typos have been addressed. You will see that "a ME-Model" remains since there is a decision how to pronounce the name of our model. We refer to it as the ME (as in 'Come along with me.')-Model.

Reviewer #3

The present paper represents a major advance in constraints-based metabolic modeling by incorporating expression of individual genes, not just a singular biomass objective function. If the key methodological issues (itemized below) can be addressed, and the claims (esp. in title and abstract) toned down to better match the results, I am supportive of publication:
1. Title: I do not see the paper as presenting a unified model of growth since it does not deal with regulation, cell division, geometry, etc. Also, looking carefully, I am not clear that it predicts different proteomes for different nutrient limitations (e.g., does it predict high GS expression in N-limitation?).

We have updated the title.

2. Abstract: The term "molecular details" is misleading. This suggests a much more precise understanding than provided here. Also, predictions of gene expression are not validated (e.g. what happens in different nutrient limitations?)

We have removed the term "molecular details".

3. General: I find the first half of the results (up to the point of the gene expression analysis) to be highly important and an excellent test of the capabilities of the model. Thereafter, the paper seems to go into a highly speculative direction that exceeds the capabilities of the model.

A few sentences have been added in the Results to reflect that the gene expression analysis is an emerging capability. With respect to global quantitative prediction of these values, we now explicitly state that the correlation is weak. Along with the changes requested by the other reviews, we hope we are no longer reaching too far beyond the means of the current model's capabilities.

4. Evolution: The comments about evolution are dubious as these may be limited to strains with low pyrE expression (pyrimidine pseduoauxotrophy).

A disclaimer has been added to the text stating that our analysis is specific to this particular strain and evolution (batch culture in M9 medium supplemented with glycerol). The sentence reads as follows: "It remains to be seen if increasing the average flux per metabolic enzyme is a general evolutionary strategy for increasing growth rate, or if it is specific to the starting strain and conditions." Our proposed systems-level mechanism for increased growth though increase in k_{eff} is consistent with relief of the pyrimidine pseudoauxotrophy. Also, Conrad et al. (2010) reported that the "pyrE gene was not observed to change in the three glycerol-adaptive mutants, so it is not immediately clear that these RNAP mutations are adaptive through effects on the pyrimidine biosynthetic pathway." Interestingly, evidence that appeared after our initial submission indicates that changes in k_{eff} are a general strategy to increase growth rate, at least in moving from nutrient limited to batch growth conditions (Valgepea et al., 2013, Mol Biosyst).

5. Fluxes: Figure 3 should also present flux changes across conditions (how well are these predicted?), and should be supported by a table that provides detailed data/model predictions. Outliers should be discussed.

Supplementary Figure 2 shows the predicted versus measured flux fold changes across conditions. We have added Supplementary Table 7 that compares fluxomic data, M-Model, and ME-Model fluxes. We have added a section to the Supplementary Information "Discussion of central carbon flux predictions" where outliers are discussed.

6. Figure 1C: Why are translation rates other than 20 AA/s not explored? It seems a modest change could better fit the data.

Decreasing the translation rate below 20 aa/s does indeed increase the predicted RNA content. However, based on equation 2 in "Supplementary Information, Hyperbolic ribosomal catalytic rate", any constant translation rate will result in no RNA produced at the limit of no growth; it is the hyperbolic rate model that can explain the non-zero intercept in Figure 1C. We added a sentence to the main text, "Derivation of constraints on molecular catalytic rates" to clarify this reasoning.

7. Methods: The optimization procedure is a hugely important aspect of this work, and should not be limited to the supplement.

We agree that this is a very important aspect of our work. We have significantly expanded the description in the supplement, and added a key sentence to the main text advertising this addition.

8. Batch optimization method: The procedure is not clear. I am an expert in the field but cannot understand it. Why is a random peptide invoked? Why are there multiple LP problems in the process? A MUCH BETTER EXPLANATION IS ESSENTIAL.

Thank you for this comment-- we agree that a clear explanation of the optimization method is essential for the paper. We have expanded the supplemental text and added a figure describing the method in the Supplementary Information under the heading of "Optimization procedure".

9. Nutrient-limited optimization method: The procedure is clear to me, but does it differentiate the importance of different enzymes in different nutrient limitations? It seems not. If so, explain how. If not, a clear disclaimer is needed.

The model presented does not differentiate the importance of different enzymes in different nutrient limitations. Further parameterization of the coupling constraints would be required for this. We have made explicit our disclaimer in the main text at the end of “Derivation of constraints on molecular catalytic rates”.

10. Janusian regime definition: Cool name, and interesting concept, but the evidence that this really exists only in C-limitation is missing. Seems like it would naturally exist also in other limitations, but that this just isn't captured in the present model.

The Janusian region is identified by the computational procedure outlined in “Supplementary Information, Optimization procedure, Computational definition and identification of growth regions”. Based on this procedure, we did not identify a Janusian region under non-carbon limitation. It may be that inclusion of substrate affinities (Km) for alternative nutrient assimilation pathways would reveal a Janusian region for non-carbon limitation (also related to comment #9). We have removed text that suggests the region only exists for carbon limitation.

2nd Editorial Decision — 03 September 2013

Thank you again for submitting your work to Molecular Systems Biology. We have now finally heard back from the two referees who accepted to evaluate the revised study. As you will see, the referees find the topic of your study of potential interest and are globally supportive. Reviewer #3 still raises issues that need to be addressed in a minor revision:
- in view of the low correlation between predicted gene expression profiles and measurements, the claim that the model can predict optimal gene expression should be considerably toned down.
- in this context, the analysis of putative mechanisms underlying the simulated gene expression appears also too speculative at this stage and this aspect should be removed. The potential for the model to enable such analysis can of course be mentioned in the Discussion section.
- similarly, the referee feels that the evolutionary part remains too speculative and we would rather prefer that a discussion about the potential for such analyses are moved to the Discussion section.

On a more editorial level, we would suggest to include the term "ME-model" in the abstract to facilitate future searches. We appreciate that you changed the title to "A constraint-based approach for computation of growth-optimizing proteomes", however it remains perhaps somewhat cryptic, in particular with regard to the fact that the model integrates metabolism and gene-expression and combines two reaction networks. Could we suggest something along the line "An integrated metabolic and gene expression model for computation of bacterial optimal growth"?

REFEREE REPORTS

Reviewer #1 :

The authors have properly addressed our comments and we are satisfied with the manuscript in its present form.

Reviewer #3 :

The present paper represents an important advance by moving beyond metabolism-specific metabolic flux models to models that include also protein expression. The title and abstract are improved. The methods are improved. My problems with the 2nd half of the results, starting on page 8, however persist. The authors readily admit that they do not predict increased expression of nitrogen assimilation genes in nitrogen limitation. Thus, the statement "the ME-model can also predict optimal gene expression in a particular nutrient environment" is patently false. This must be fixed. In terms of the rest, to me, it is acceptable to show the general trends in figure 4 and figure 5b, but to proceed to analyze the underlying transcription factors becomes a case of layering computational analysis upon computational analysis when experimentation would do just fine (and indeed has already been done). In the evolutionary part, in figure 6, I see no genuine evidence that the authors are correct and I personally disfavor such speculation in a first-rate journal.

2nd Revision - authors' response 04 September 2013

Point-by-Point Response (minor revision round)

We have revised our manuscript based on the reviewer/editor comments and addressed the issues that were raised.

Below, please find our detailed response to the comments of the reviewers. We thank the reviewers for their thoughtful comments and questions, and hope the revised manuscript will now be accepted for publication.

Thank you again for submitting your work to Molecular Systems Biology. We have now finally heard back from the two referees who accepted to evaluate the revised study. As you will see, the referees find the topic of your study of potential interest and are globally supportive. Reviewer #3 still raises issues that need to be addressed in a minor revision:
- in view of the low correlation between predicted gene expression profiles and measurements, the claim that the model can predict optimal gene expression should be considerably toned down.

This claim was toned down considerably throughout.

- in this context, the analysis of putative mechanisms underlying the simulated gene expression appears also too speculative at this stage and this aspect should be removed. The potential for the model to enable such analysis can of course be mentioned in the Discussion section.

This aspect was removed from the Results as requested. As a result, Figure 4C and Figures 5D-5E from MSB-13-4501R are not part of the newly revised manuscript. The potential for the model to enable such analysis is now mentioned in the Discussion.

- similarly, the referee feels that the evolutionary part remains too speculative and we would rather prefer that a discussion about the potential for such analyses are moved to the Discussion section.

We implemented this suggestion exactly as stated above. As a result, Figure 6 from MSB-13-4501R is not part of the newly revised manuscript. The potential for the model to enable such analysis is now mentioned in the Discussion.

On a more editorial level, we would suggest to include the term "ME-model" in the abstract to facilitate future searches. We appreciate that you changed the title to "A constraint-based approach for computation of growth-optimizing proteomes", however it remains perhaps somewhat cryptic, in particular with regard to the fact that the model integrates metabolism and gene-expression and combines two reaction networks. Could we suggest something along the line "An integrated metabolic and gene expression model for computation of bacterial optimal growth"?

The term "ME-Model" was added to the Abstract and the title was revised along these lines.

Reviewer #1:
The authors have properly addressed our comments and we are satisfied with the manuscript in its present form.

:)

Reviewer #3:

The present paper represents an important advance by moving beyond metabolism-specific metabolic flux models to models that include also protein expression. The title and abstract are improved. The methods are improved.

My problems with the 2nd half of the results, starting on page 8, however persist. The authors readily admit that they do not predict increased expression of nitrogen assimilation genes in nitrogen limitation. Thus, the statement "the ME-model can also predict optimal gene expression in a particular nutrient environment" is patently false. This must be fixed. In terms of the rest, to me, it is acceptable to show the general trends in figure 4 and figure 5b, but to proceed to analyze the underlying transcription factors becomes a case of layering computational analysis upon computational analysis when experimentation would do just fine (and indeed has already been done). In the evolutionary part, in figure 6, I see no genuine evidence that the authors are correct and I personally disfavor such speculation in a first-rate journal.

Thank you for these comments. As some of the results presented in Figure 4C, Figures 5D-5E, and Figure 6 were underdeveloped, they were completely removed. Only the potential for such analysis was mentioned in the discussion section.